基于反向选择的地震预测方法 *

吴晶晶，梁意文，谭成予，周　雯

(武汉大学 计算机学院，武汉 430072)

摘　要：针对大地震历史数据缺乏导致的大地震预测准确率低问题，提出一种基于反向选择的地震预测方法。采用可变实值反向选择算法生成成熟检测器，用于预测地震是否发生。由于反向选择在训练过程中无须非我数据，可减小大地震数据缺乏对训练效果的影响。实验采用四川省历史地震数据，对一个月内是否发生 5.0 级及以上地震进行预测。与传统机器学习算法进行对比，结果表明反向选择算法具有更好的预测效果。

关键词：反向选择；地震预测；地震活动性指标；神经网络

中图分类号：TP311　　doi: 10.3969/j.issn.1001-3695.2017.10.0783

Method of earthquake prediction based on negative selection

Wu Jingjing, Liang Yiwen, Tan Chengyu, Zhou Wen

(*School of Computer*, *Wuhan University*, *Wuhan* 430072, *China*)

Abstract: For the low accuracy of earthquake prediction caused by the lack of large earthquake data, this paper proposed an earthquake prediction method based on negative selection. In this method, it used the variable real valued negative selection algorithm to generate mature detectors which used to predict whether an earthquake occured. Due to the absence of non-self data sets in the negative selection training, it could be reduced the impact of lacking large earthquake data on the training effect. It used the historical earthquake data of Sichuan province to predict whether the magnitude 5.0 and above earthquakes occurred in Sichuan within one month. Compared with the traditional machine learning algorithm, the results show that the negative selection has better prediction effect.

Key words: negative selection; earthquake prediction; seismicity indicator; neural network

0　引言

地震预测是对未来地震事件发生的具体地点、震级和时间进行准确预测的过程[1]。通过专家们长期观测研究和经验积累发现，地震三要素（地震发生时间、震级、震中位置）与特征指标之间存在着内在联系[2]，利用地震特征指标的变化预测是一种常用的地震预测方法。

随着人工智能机器学习的发展，国内外研究学者尝试将相关算法应用在地震预测中，并取得了较好的预测效果。然而，由于大地震的发生是一种小概率事件，发生时间间隔较长，利用现代观测设备观测以来发生的大地震数量相对较少，可用于研究的数据缺乏，影响通过历史地震数据进行的大地震预测准确率[1]。而神经网络、支持向量机等机器学习方法需要大量训练数据集，现有地震预测方法在对大地震的预测上存在准确度不高问题。

计算机免疫学的反向选择算法[3]模拟了生物免疫系统中自我—非我（self-nonself）识别原理。该算法借鉴生物体内的未成熟 T 淋巴细胞在成熟过程中会进行反向选择的过程。首先确定自我集合；然后随机产生检测器，通过删除识别到自我的检测器，而保留识别非我的检测器。其特点是不需要先验知识，只需使用有限数量的自我数据作为训练集，就可以生成大量非我检测器，用于检测非我[3]。由于反向选择能识别自我和非我，其常被用于二分类问题。本文预测 5.0 及以上地震是否发生，本质上可看做二分类问题[4]，可使用反向选择算法识别地震发生（非我）和地震不发生（自我）。由于反向选择训练过程中无须非我数据集（5.0 及以上地震发生），可减小传统机器学习方法缺少大地震数据导致的预测准确度不高问题。

本文引入计算机免疫学的反向选择算法建立地震预测模型。利用反向选择的自我—非我识别原理，将预定义阈值及以上震级的地震未发生作为自我，预定义阈值及以上震级的地震发生作为非我。由于只使用了自我数据作为训练集，减小了大地震数据缺乏对训练效果的影响，达到提高大地震预测准确度的目

收稿日期：2017-10-28；**修回日期：**2017-12-27　　**基金项目：**国家自然科学基金资助项目（61170306）

作者简介：吴晶晶（1992-），女，湖北随州人，硕士研究生，主要研究方向为计算机免疫学、神经网络（whu_wujingjing@163.com）；梁意文（1962-），男，教授，博导，主要研究方向为计算机免疫、数据起源；谭成予（1970-），女，副教授，硕导，主要研究方向为计算机免疫、数据起源；周雯（1989-），女，博士，主要研究方向为计算机免疫学.

的。本文实验部分以 5.0 级震级作为阈值。以下描述中均以 5.0 级作为预测阈值。

1　相关工作

地震的发生被认为是一种随机的、高度非线性的现象，现有条件下很难完全准确地预测地震发生的时间、位置和震级。1939 年，Ishimoto 和 Iida 首次进行了地震参数预测，提出震级与地震数量之间的关联关系[5]。1999 年，Sykes 等人[6]认为对于非常活跃的地震指标变化，大地震预测是可能的。随着计算机相关方法在地震预测中的应用，地震预测研究取得了进步。常用于地震预测的计算机方法主要是基于人工智能的方法，采用机器学习中的监督式学习和非监督式学习方法进行地震预测[7]。

监督式学习方法包括人工神经网络[8~13]、支持向量机[7,14]、贝叶斯网络[7]、决策树[7]等。其中人工神经网络是地震预测中使用最多的机器学习算法。2002 年，Negarestani 等人[8]使用反向传播神经网络来检测地震引起的氡浓度的异常行为。2004 年，Liu 等人[15]通过使用径向基函数神经网络预测了中国地区地震。2007 年，Panakkat 等人[9]引入了一种新的地震预测方法，采用从历史地震事件中通过数学计算的八个地震指标，使用三种不同的神经网络方法预测地震。2015 年，Asencio-Cortés 等人[7]分别使用四种机器学习算法（KNN、支持向量机、贝叶斯网络、决策树算法）预测了东京地震，实验结果表明神经网络具有更好的预测效果。

非监督式学习算法包括关联规则[16,17]、K-means 算法[18]等。2011 年，Martí nezálvarez 等人[16]使用计算智能技术用于预测地震，通过提取定量关联规则和回归技术，用于建立地震时间数据的行为模型。2014 年，Florido 等人[18]使用 K-means 算法，通过对智利地震时间序列分析，预测了智利大于 4.4 级地震发生的概率。2015 年，Ikram 等人[17]开发了一个基于关联规则和谓词逻辑的专家系统用于地震预测，使用历史地震数据，通过关联规则挖掘地震数据。

现有用于地震预测的机器学习算法大多需要大量训练数据集，而大地震数据缺乏，用于预测的大地震样本数据较少，导致预测准确率不高。

1994 年，Forrest 等人[3]受生物免疫系统 T 细胞成熟过程中的反向选择原理启发，提出了反向选择算法。随后，Zhou 和 Dasgupta [19,20]等人对该模型进行了扩展。该算法模拟了免疫系统中 T 细胞成熟的反向选择过程中识别自我和非我的免疫耐受过程。由于反向选择的自我—非我识别能力，其被成功应用于故障检测、异常检测、分类等问题。本文预测的是 5.0 级及以上地震发生或不发生，可看做是一个二分类问题。可将反向选择应用于地震预测，将 5.0 级及以上地震不发生作为自我，5.0 级及以上地震发生作为非我。反向选择算法在训练过程中只需使用自我数据，即 5.0 级及以上地震不发生作为训练数据集，因此可减少大地震数据缺乏对训练效果的影响，从而提高预测准确度。

综上所述，地震预测模型的选取对预测效果至关重要。本文提出一种新的地震预测方法，选取文献[9]使用的八个指标作为特征向量，使用反向选择建立地震预测模型。由于反向选择算法的检测器生成过程只需使用自我数据集，可用来减小神经网络等方法缺乏大地震数据导致的预测准确度不高问题。

2　地震预测方法

2.1　地震活动性指标

文献[9]提出使用八个通过数学定义的地震参数即地震活动性指标来评估一个地区的地震发生潜能，为使用计算机方法预测地震提供了新途径。文献[7~21]均使用了这八个地震活动性指标。

本文使用上述八个地震活动性指标，每余个指标使用 n（如 100）个历史地震数据通过数学计算得出。指标描述及计算公式如表 1 所示。其中指标 b 、η 、ΔM 基于 Gutenberg-Richter 逆幂律，指标 μ 、c 基于特征地震震级分布，指标 T 、M_{mean} 、$dE^{1/2}$ 来源于地震预测研究经验得出的结论。特征地震事件表示 n 个地震事件中 5.0 级及以上震级的地震事件。

表 1　地震活动性指标描述

地震活动性指标	描述	计算公式
T	n 个地震事件时间间隔	$T = t_n - t_1$
M_{mean}	n 个地震时间的平均震级	$M_{\mathrm{mean}} = \sum E^{1/2} / T$
$dE^{1/2}$	E 为地震释放的能量；$dE^{1/2}$ 表示地震能量释放的平方根率	$\mathrm{d}E^{1/2} = \sum E^{1/2} / T$
b	地震频率的对数与震级曲线的斜率	$\mathrm{b} = \frac{(n\sum(M_i \lg N_i) - \sum M_i \sum \lg N_i)}{((\sum M_i)^2 - n\sum M_i^{\ 2})}$
η	均方偏差，基于 Gutenberg –Richter 逆幂律的回归线的均方差总和，反映震级与频率之间的关系	$\mathrm{a} = \sum(\log_{10} N_i + bM_i) / n$ $\eta = \frac{\sum(\log_{10} N_i - (a - bM_i))^2}{(n-1)}$
ΔM	观察到的最大震级与预期的最大震级之间的差异	$M_{\max,expected} = a / b$ $\Delta M = M_{\max,observed} - M_{\max,expected}$
μ	特征事件之间的平均时间间隔	$\mu = \sum(t_{i characteristic}) / n_{characteristic}$
c	特征事件之间的平均时间的变异系数	c = 观测到的时间的标准偏差 / μ

2.2　反向选择算法

反向选择的目的是产生对于自我细胞的耐受能力，使得免疫系统具备检测未知抗原并且不与自我细胞发生反应的能力[3]。

本文采用可变实值反向选择算法[19,22]，算法输入为 self 集合，每个 self 由一个 n 维向量表示，算法生成的检测器（抗体）覆盖 nonself 空间。主要是使用一个 n 维向量定义检测器，用实值描述检测器的识别检测半径。一个检测器可被认为是超球体。

检测器是否检测抗原取决于检测器与抗原之间的距离（使用欧几里得距离公式计算）和检测器半径。如果距离小于 self 半径，说明两个超球体相交，即表明检测器能识别 self，则该检测器被淘汰；反之，则将检测器与 self 之间的距离作为半径。成熟检测器半径不同，则对于大范围 nonself 空间可以用很少的识别器覆盖，同时对于 self 周围较少的 nonself 空间，可用半径较小的检测器覆盖。成熟检测器生成算法流程如下：

```
初始化检测器数 num、self 半径 rs
初始化检测器集合 D//D 为空
输入训练集 S//自我抗原集合
初始化最大距离 distance
do{
  随机生成 n 维向量的随机检测器 d
  for 训练集 S 中每个自我抗原 s
  {
     计算随机检测器 d 与自我的欧几里得距离 dis
     If(dis<distance) distance = dis;
  }
  if(rs<distance)
  {
     rd = distance-rs//检测器半径
     将向量<d,rd>加入检成熟检测器集合 D 中
  }
}while(D 的数量达到 num)
```

2.3 基于反向选择的地震预测方法

本文采用通过历史数据计算得出的八个地震活动性指标作为单个样本数据。将样本数据根据 5.0 级及以上地震不发生和发生分为 self 和 nonself 集合。其中 5.0 级及以上地震发生标记为 1，5.0 级及以上地震不发生标记为 0。使用反向选择算法作为预测模型。主要步骤如下：

a)数据预处理。获取历史地震数据，根据 3.1 节的数学公式计算地震活动性指标，形成 8 维实数向量。

b)定义 self 集。5.0 级及以上地震不发生；定义 nonself 集：5.0 级及以上地震发生。

定义 *ALL* 表示所有可能的预测结果分类集合。由于预测的是 5.0 级及以上地震是否发生，所以预测结果分类可以分为两类：预测 5.0 级及以上地震发生；预测 5.0 级及以上地震不发生。因此，可将自我集合 *SELF* 定义为 5.0 级及以上地震不发生，非我集合 *NONSELF* 定义为 5.0 级及以上地震发生。三者之间满足以下关系：

$$SELF \cup NONSELF = ALL \tag{1}$$

$$SELF \cap NONSELF = \varnothing \tag{2}$$

c) 检测器生成。通过可变实值反向选择算法生成 n 个成熟检测器。

d)地震预测。使用成熟检测器集合匹配输入的抗原向量。检测抗原为自我表示 5.0 级及以上地震不发生，输出 0；检测抗原为非我表示 5.0 级及以上地震发生，输出 1。

地震预测模型可表示为三元组：$\Sigma_{ADS}=(IN_{ADS}, OUT_{ADS}, F_{ADS})$。其中：$IN_{ADS}$ 表示预测模型输入，有 $IN_{ADS}\in ALL$；OUT_{ADS} 表示预测模型输出，输出分为两种预测结果，分别用 0 和 1 表示；F_{ADS} 表示输入与输出的函数关系，满足如下关系：

$$OUT_{ADS}=F_{ADS}(IN_{ADS})=\begin{cases}1, if & IN_{ADS}\in SELF\\ 0, if & IN_{ADS}\in NON-SELF\end{cases} \tag{3}$$

模型分为检测器生成阶段和地震预测阶段，如图 1、 2 所示。

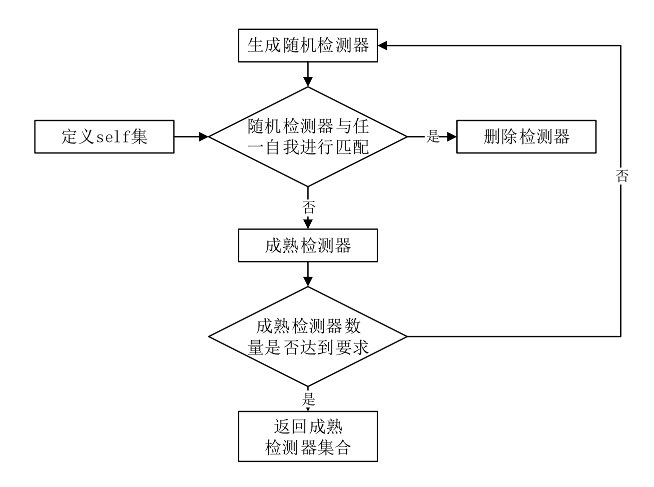


图 1 检测器生成阶段

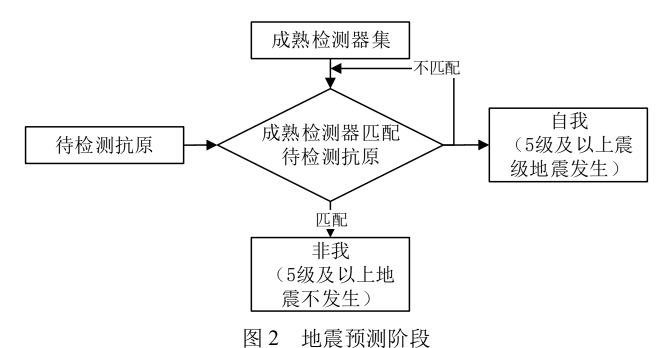


图 2 地震预测阶段

3 实验与结果分析

本文设计了基于反向选择的地震预测实验，预测四川省一个月内是否会发生预定义阈值及以上震级的地震。本文实验将预测震级阈值设置为 5.0 级。实验目的在于验证反向选择算法对于地震预测的有效性和准确性，对比实验采用两种机器学习算法：神经网络、支持向量机算法。实验分为数据预处理、地震预测、准确度评估三部分。

3.1 数据预处理

数据采用四川省 1990 年 1 月到 2016 年 12 月的历史地震数据（包括日期、时间、经度、纬度、深度、震级），数据的来源为“中国地震台网”。参考文献[9]等地震预测文献研究成果，本文只使用 4.0 级及以上历史地震数据。历史地震数据共 928 条，共预测 1991 年 10 月到 2016 年 12 月共 305 个月的地震。表 2 为 1991 年 4.0 级及以上历史地震记录。预测时间为一个

月，使用预测月份之前的 100 条地震数据计算出接下来一个月的八个地震指标，得出 305 组数据集。将前 204 组地震指标中的 198 组自我数据集挑选出来作为训练集，后 101 组地震指标作为测试集。

表 2 1991 年 4.0 级及以上原始数据

时间	纬度	经度	深度	震级
1991/2/18 9:06:22	31.66	102.53	20	5.1
1991/4/11 20:32:18	27.23	101.13	8	4.7
1991/10/28 18:43:24	31.64	102.48	8	4
1991/11/3 0:02:26	28.31	104.03	33	4.4
1991/12/26 13:24:17	30.97	99.62	18	4.4
1991/12/31 21:14:15	30.78	99.6	11	4.3

针对实验中采用的样本数据集，由于用于代表自我—非我的数据样本是八个地震指标组成的 8 维向量，由预测月之前 100 条地震数据计算得出，并非未经处理的历史地震数据，所以尽管计算地震指标的 100 条数据只包含了 4.0 级及以上震级地震，不包含 4.0 级以下及未发生地震的历史数据，缺少的数据不属于本文定义的自我非我，对本文数据样本是否能代表自我非我空间不产生影响。

3.2 评估指标

本文参考文献[9]提到的评估标准评估方法性能，地震预测结果与实际观察结果矩阵如表 3 所示。

表 3 地震预测结果与实际观察结果矩阵

	观测到地震发生	观测到地震未发生
预测地震发生	N_{pc}	N_{ni}
预测地震不发生	N_{pi}	N_{nc}

根据 N_{pc}、N_{ni}、N_{pi}，分别计算 POD、FAR 和 R 技能得分，具体如下所示：

a) 检测概率（POD）。POD 越高，表示预测正确率越高。

$$POD = \frac{N_{pc}}{N_{pc} + N_{pi}} \tag{4}$$

b) 误报率（FAR）。FAR 越高，表示预测的误报率越大。

$$FAR = \frac{N_{ni}}{N_{ni} + N_{nc}} \tag{5}$$

c) R 技能得分。R=1 表示全部预测到，R=0 表示全部没预测到，即 R 得分越大，预测效果越好。

$$R = POD - FAR = \frac{N_{pc}}{N_{pc} + N_{pi}} - \frac{N_{ni}}{N_{ni} + N_{nc}} \tag{6}$$

3.3 实验结果分析

实验分为三组进行，使用本文提出的基于反向选择的地震预测方法和两种机器学习方法（神经网络、支持向量机）作对比。分别对 101 组测试集进行预测，预测输出两种分类：0（预测地震不发生）、1（预测地震发生）。根据预测结果统计出 N_{pc}、N_{ni}、N_{pi}、N_{nc}，并分别计算出检测概率 POD、误报率 FAR 和 R 技能得分这三个评估指标。实验结果如表 4 所示。其中 R 值由 Xu[24]提出，用于表示地震预测能力评分。R 值为成功率减去误报率，值为 1 时，表示完全预测正确；值为 0 时，表明预测无任何意义；值为负值时，表示完全预测错误。因此,当 R 值大于 0 时，表示所作的预测成功率高于误报率，即预测是起到一定作用的，R 值越大，则地震预测效果越好。因此使用 R 来评估预测模型预测效果。

表 4 实验结果对比

评估指标	反向选择	神经网络	支持向量机
N_{pc}	12	10	3
N_{ni}	7	9	16
N_{pi}	4	6	13
N_{nc}	78	74	69
POD	0.7500	0.6250	0.1875
FAR	0.0824	0.1084	0.1882
R	0.6676	0.5166	-0.0007

由于神经网络、支持向量机在训练过程中需要大量训练数据集，而用于预测的大地震数据较少，会对训练结果产生影响，降低预测准确率。反向选择算法在训练过程中无须非我数据集，可减小大地震数据缺乏对训练效果的影响，从而提高预测准确度。从实验结果可以看出，支持向量机 R 值为负，预测效果较差。反向选择算法相对于神经网络 R 值提高 15.1%，误报率降低 2.6%。以上分析结果证明，反向选择算法在地震预测上相比传统机器学习算法（神经网络和支持向量机）具有更好的预测能力，能减小大地震数据缺乏对训练效果的影响，提高预测准确率。

4 结束语

本文提出基于反向选择的地震预测方法，对四川省地区地震进行预测。实验表明，反向选择算法较其他两种机器学习算法在地震预测方面具有更好的效果。尽管如此，反向选择算法在地震预测的准确度仍然不够理想，无法达到准确预测的目的。下一步将继续完善预测模型。本文预测的是 5.0 级及以上震级地震是否发生，未预测出具体震级，下一步将考虑如何对具体震级进行预测。

参考文献：

[1] 陈运泰. 地震预测：回顾与展望 [J]. 中国科学：地球科学 D 辑：地球科学, 2009, 39 (12): 1633-1658.

[2] Panakkat A, Adeli H. Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators [J]. Computer-Aided Civil and Infrastructure Engineering, 2009, 24 (4): 280–292.

[3] Forrest S, Perelson A S, Allen L, *et al.* Self-nonself discrimination in a computer [C]// Proc of IEEE Computer Society Symposium on Research in

Security and Privacy. 1994: 202-212.

[4] Fernández-Gómez M, Asencio-Cortés G, Troncoso A, *et al.* Large earthquake magnitude prediction in chile with imbalanced classifiers and ensemble learning [J]. Applied Sciences, 2017, 7 (6): 625.

[5] Panakkat A, Adeli H. Recent efforts in earthquake prediction (1990–2007) [J]. Natural Hazards Review, 2008, 9 (2): 70-80.

[6] Sykes L R, Shaw B E, Scholz C H. Rethinking earthquake prediction [J]. Pure and Applied Geophysics, 1999, 155 (2): 207-232.

[7] Asencio-Cortés G, Martínez-Álvarez F, Troncoso A, *et al.* Medium-large earthquake magnitude prediction in Tokyo with artificial neural networks [J]. Neural Computing & Applications, 2017, 28 (5): 1043-1055.

[8] Negarestani A, Setayeshi S, Ghannadimaragheh M, *et al.* Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction [J]. Journal of Environmental Radioactivity, 2002, 62 (3): 225.

[9] Panakkat A, Adeli H. Neural network models for earthquake magnitude prediction using multiple seismicity indicators [J]. International Journal of Neural Systems, 2007, 17 (1): 13-33.

[10] Adeli H, Panakkat A. A probabilistic neural network for earthquake magnitude prediction [J]. Neural Networks the Official Journal of the International Neural Network Society, 2009, 22 (7): 1018-1024.

[11] Panakkat A, Adeli H. Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators [J]. Computer-Aided Civil and Infrastructure Engineering, 2009, 24 (4): 280-292.

[12] Asim K M, Martínez-Álvarez F, Basit A, *et al.* Earthquake magnitude prediction in Hindukush region using machine learning techniques [J]. Natural Hazards, 2017, 85 (1): 1-16.

[13] Reyes J, Morales-Esteban A, Martínez-Álvarez F. Neural networks to predict earthquakes in Chile [J]. Applied Soft Computing, 2013, 13 (2): 1314-1328.

[14] Asencio-Cortés G, Scitovski S, Scitovski R, *et al.* Temporal analysis of croatian seismogenic zones to improve earthquake magnitude prediction [J]. Earth Science Informatics, 2017, 10 (3): 303-320.

[15] Liu Yue, Wang Yuan, Li Yuan, *et al.* Earthquake prediction by RBF neural network ensemble [C]// Proc of International Symposium on Neural Networks. 2004: 962-969.

[16] Martínezálvarez F, Troncoso A, Moralesesteban A, *et al.* Computational intelligence techniques for predicting earthquakes [C]// Lecture Notes in Computer Science. 2011: 287-294.

[17] Ikram A, Qamar U. Developing an expert system based on association rules and predicate logic for earthquake prediction [J]. Knowledge-Based Systems, 2015, 75 (C): 87-103.

[18] Florido E, Martínez-Álvarez F, Aznarte J L, *et al.* Discovering patterns preceding earthquakes in Chilen time series [C]// Proc of International Work-Conference on Time. 2014.

[19] Zhou J, Dasgupta D. Real-valued negative selection algorithm with variable-sized detectors [C]// Lecture Notes in Computer Science. 2004: 287-298.

[20] Dasgupta D, Krishnakumar K, Wong D, *et al.* Negative selection algorithm for aircraft fault detection [C]// Proc of the 3rd International Conference Artificial Immune Systems. 2004: 1-13.

[21] Asenciocortés G, Martínezálvarez F, Moralesesteban A, *et al.* Improving earthquake prediction with principal component analysis: application to chiles [C]// Lecture Notes in Computer Science. 2015.

[22] Alizadeh E, Meskin N, Khorasani K. A negative selection immune system inspired methodology for fault diagnosis of wind turbines [J]. IEEE Trans on Cybernetics, 2017, 47 (11): 3799-3813.

[23] Asim K M, Martínez-Álvarez F, Basit A, *et al.* Earthquake magnitude prediction in Hindukush region using machine learning techniques [J]. Natural Hazards, 2017, 85 (1): 1-16.

[24] Xu Shaoxie. Method of prediction of earthquake by using seismisity [J]. Acta Seismologica Sinica, 1993, 6 (2): 517-533.